



ELSEVIER

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

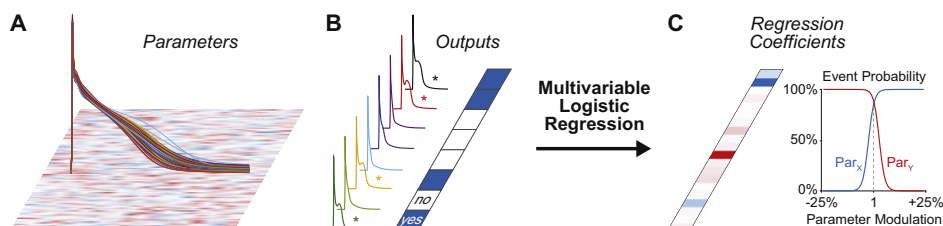
Logistic regression analysis of populations of electrophysiological models to assess proarrhythmic risk



Stefano Morotti, Eleonora Grandi*

Department of Pharmacology, University of California Davis, Davis, CA, USA

GRAPHICAL ABSTRACT



ABSTRACT

Population-based computational approaches have been developed in recent years and helped to gain insight into arrhythmia mechanisms, and intra- and inter-patient variability (e.g., in drug responses). Here, we illustrate the use of multivariable logistic regression to analyze the factors that enhance or reduce the susceptibility to cellular arrhythmogenic events. As an example, we generate 1000 model variants by randomly modifying ionic conductances and maximal rates of ion transports in our atrial myocyte model and simulate an arrhythmia-provoking protocol that enhances early afterdepolarization (EAD) proclivity. We then treat EAD occurrence as a categorical, yes or no variable, and perform logistic regression to relate perturbations in model parameters to the presence/absence of EADs. We find that EAD formation is sensitive to the conductance of the voltage-gated Na⁺, the acetylcholine-sensitive and ultra-rapid K⁺ channels, and the Na⁺/Ca²⁺ exchange current, which matches our mechanistic understanding of the process and preliminary sensitivity analysis.

The described technique:

- allows investigating the factors underlying dichotomous outcomes, and is therefore a useful tool to improve our understanding of arrhythmic risk;
- is valid for analyzing both deterministic and stochastic models, and various phenomena (e.g., delayed afterdepolarizations and Ca²⁺ sparks);

* Corresponding author at: Department of Pharmacology, University of California Davis, 451 Health Sciences Drive, GBSF rm 3502, Davis, CA 95616, USA.

E-mail address: ele.grandi@gmail.com (E. Grandi).

<http://dx.doi.org/10.1016/j.mex.2016.12.002>

2215-0161/© 2016 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

- is computationally more efficient than one-at-a-time parameter sensitivity analysis.

© 2016 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

ARTICLE INFO

Method name: Logistic regression analysis of populations of electrophysiological models

Keywords: Mathematical modeling, Cardiac arrhythmia, Sensitivity analysis, Logistic regression

Article history: Available online 23 December 2016

Method details

Population-based approaches have been widely adopted in the recent years in the field of computational cardiac electrophysiology [1–8]. The classical analysis performed with a single action potential (AP) model has been extended to the study of the properties emerging in a broad group of models (called “population”). Each individual in the population is obtained from the same parent model (called “baseline”) by randomly perturbing several parameters. These differences in model parameterization reflect the natural variability among cells in a tissue (intra-subject variability) or among different individuals (inter-subject variability). Thus, by investigating a more realistic and comprehensive scenario, this approach allows obtaining new insights in several aspects of myocyte function in health and disease, including arrhythmia mechanisms and responses to drugs [9]. Moreover, the assessment of sensitivity of model predictions to parameter perturbations ensures enhanced robustness of the results of computational studies [10]. However, the choice of appropriate mathematical and statistical tools is fundamental when performing sensitivity analysis on a population of hundreds or thousands AP models. Here, we illustrate the use of multivariable logistic regression to analyze the factors underlying the occurrence of a dichotomous outcome (such as the development of cellular arrhythmogenic events).

Explanation of the methodology

Regression analysis is a statistical process used to estimate relationships among variables in a system [11,12]. It is usually applied to correlate modifications in independent variables (called “predictors”) to the consequent effect on dependent variables. The result of regression analysis is a function of the independent variables (called “regression function”), which is used to predict the values of the dependent variables (e.g., to estimate a certain outcome after changes in fixed system properties). More specifically, the result of this analysis is the set of regression coefficients used in the regression function. The general principles of regression analysis are formalized in the following equation:

$$\mathbf{X} * \mathbf{B} = \mathbf{Y} \approx \tilde{\mathbf{Y}}$$

where \mathbf{X} and \mathbf{Y} are the matrices of the independent and dependent variables, respectively. Based on \mathbf{X} and \mathbf{Y} , this technique allows the determination of the regression coefficients (in the matrix \mathbf{B}), which can be used to estimate \mathbf{Y} ($\tilde{\mathbf{Y}}$) given \mathbf{X} .

In the field of computational cardiac electrophysiology, sensitivity analysis of population of AP models is usually performed to assess how variations in model parameters (e.g., conductances of ion channels) affect model outputs, such as AP duration (APD) or Ca^{2+} transient amplitude. In this case, when the outputs of interest are continuous variables, linear multivariable regression analysis can be performed, and the resulting regression function allows predicting how these outputs change (in a continuous range) upon modifications in model parameters [1]. In a population of n_m AP models, obtained by varying n_p parameters in a baseline model, the size of the matrix of the predictors \mathbf{X} is $n_m \times n_p$ (rows \times columns). The matrix \mathbf{Y} reports different outputs (n_o) identified in each model of the population (size $n_m \times n_o$). In general, the matrix \mathbf{B} contains a coefficient describing the effect of perturbation in each parameter on each output (size $n_p \times n_o$).

A different statistical process, the multivariable logistic regression, can be used instead when the outputs of interest are described as Boolean variables (i.e., yes/no, true/false, presence/absence) [2,8]. This analysis allows estimating how perturbations in model parameters affect the probability that a certain dichotomous outcome will occur. A typical example (described in the next section) is the occurrence of cellular arrhythmogenic events such as early afterdepolarizations (EADs). The regression function used here is the logistic curve, an S-shape (sigmoidal) curve with values between 0 and 1. These extremes represent, respectively, low and high probabilities of occurrence of a certain event.

Logistic regression analysis correlates perturbations to model parameters (used to determine the matrix of the predictors **X** as described below) to the values in **Y**, which consist of binary variables corresponding to whether or not the outcome of interest has occurred. The resulting matrix **B** contains a set of $n_p + 1$ regression coefficients (array **b**) for each output studied (Fig. 1). Each set includes n_p coefficients (b_1 – b_{np}) corresponding to the effect induced by perturbations in the n_p parameters, and an additional coefficient b_0 (called “intercept term”). In the generic model configuration (described by the array **x**), values in **b** are used to assess the probability P of occurrence of the desired output as follows:

$$P(\mathbf{x}, \mathbf{b}) = 1 / (1 + \exp(-(\mathbf{b}_0 + \mathbf{x}_1 * \mathbf{b}_1 + \mathbf{x}_2 * \mathbf{b}_2 + \dots + \mathbf{x}_{np} * \mathbf{b}_{np})))$$

In the next section we describe how logistic regression is applied to the study of EAD formation and how its results can be interpreted. We will also explain how this analysis is implemented with MATLAB (The MathWorks, Natick, MA, USA). Our codes, generated with the MATLAB version R2016a,

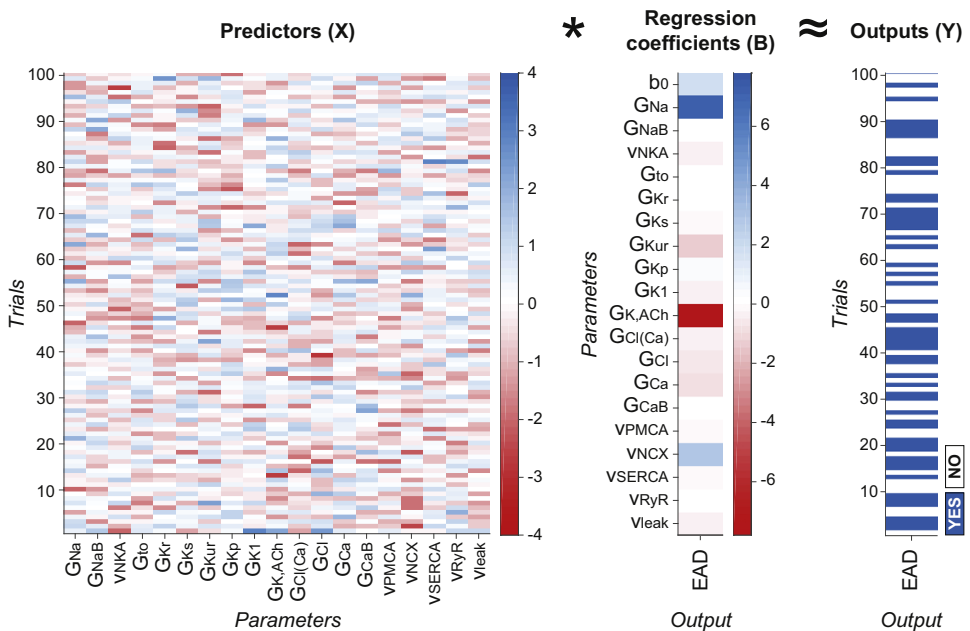


Fig. 1. Schematic representation of predictor, coefficient, and output matrices in logistic regression analysis. In the example presented in this paper, the matrix of the predictors **X** contains information about the scale factors used to generate a population of 1000 AP models (only 100 cases are shown here for the sake of simplicity) by perturbing the baseline value of 19 parameters (defined in Table 1). The same proarrhythmic protocol is simulated with all the models in the population, and the presence of EADs is assessed in each case, thereby obtaining the binary elements of the matrix of the outputs **Y**. The result of logistic regression analysis is the matrix of the coefficients **B**, used to estimate the effect of perturbations in model parameters on the probability of arrhythmogenesis. Here both **Y** and **B** are arrays because only one output (i.e., EAD development) is evaluated in this example. Values in **X** (obtained from the scale factors via log-transformation and conversion into z-scores, see text) and **B** (regression coefficients, shown also in Fig. 4) are dimensionless quantities, while elements of **Y** are Boolean variables.

are freely available for download at the following webpages: somapp.ucdmc.ucdavis.edu/Pharmacology/bers & elegrandi.wixsite.com/grandilab/downloads.

Example of application

An application of the described methodology is the evaluation of the factors influencing arrhythmia proclivity. In the example presented here, we apply logistic regression to study the mechanisms responsible for the development of EADs, treated as a Boolean variable.

In a recent computational study [13], we used our established model of AP and Ca^{2+} handling in the human atrial myocyte [14] to describe a novel mechanism for phase-3 EADs. In canine atrial preparations, Burashnikov and Antzelevitch have shown that phase-3 EADs arise after termination of atrial fibrillation (AF), and are the trigger for re-induction of fibrillation [15]. Analogous triggered phenomena have been described in canine pulmonary veins with autonomic nerve stimulation [16], or combined acetylcholine (ACh) and epinephrine challenge [17]. These protocols mimic simultaneous sympathovagal discharges that commonly precede AF onset [18–20]. Experimental studies have shown that initiation of these EADs occurs secondary to APD shortening by ACh (via ACh-sensitive K^+ current, $I_{\text{K,ACh}}$), potentiated sarcoplasmic reticulum (SR) Ca^{2+} release, and enhanced $\text{Na}^+/\text{Ca}^{2+}$ exchange (NCX) [16]. We demonstrated that these allow for cardiac Na^+ current (I_{Na}) recovery and non-equilibrium reactivation, which ultimately drives the EADs [13]. Based on data from models of acute AF, and data describing the early electrophysiological changes in AF [21], we speculated that EADs are a likely source of ectopy early in disease progression (e.g., paroxysmal AF), particularly in the pulmonary veins, where the short AP and late AP plateau (at negative membrane potential) induced by autonomic challenge can facilitate non-equilibrium I_{Na} reactivation. The same phenomenon was previously observed in mouse ventricular myocytes [22].

The described arrhythmogenic process involves many elements that are important regulators of myocyte electrophysiology and Ca^{2+} handling (such as I_{Na} , NCX and SERCA activity, and cholinergic and adrenergic signaling cascades). Thus, the use of logistic regression analysis is particularly useful in this context to assess the impact of each ion current or transporter on arrhythmia mechanisms, thereby providing potentially useful insight for the identification of therapeutic targets.

Here we use our published model [13] as a baseline to build a population of 1000 elements ($n_m = 1000$) by randomly modifying 19 parameters ($n_p = 19$) corresponding to conductances (G) of ion currents and maximal rates (v) of ion transports (see parameter definition in Table 1). We follow the approach described by Sobie [1]: for each individual, we modify the baseline value of each parameter

Table 1
Definition of the perturbed model parameters.

Parameter	Definition
G_{Na}	Conductance of the Na^+ current
G_{NaB}	Conductance of the background Na^+ current
V_{NKA}	Rate of the Na^+/K^+ pump
G_{to}	Conductance of the transient outward K^+ current
G_{Kr}	Conductance of the rapidly activating K^+ current
G_{Ks}	Conductance of the slowly activating K^+ current
G_{Kur}	Conductance of the ultra-rapidly activating K^+ current
G_{Kp}	Conductance of the plateau K^+ current
G_{K1}	Conductance of the inward rectifier K^+ current
$G_{\text{K,ACh}}$	Conductance of the acetylcholine-sensitive K^+ current
$G_{\text{Cl(Ca)}}$	Conductance of the Ca^{2+} -dependent Cl^- current
G_{ClB}	Conductance of the background Cl^- current
G_{Ca}	Conductance of the L-type Ca^{2+} current
G_{CaB}	Conductance of the background Ca^{2+} current
V_{PMCA}	Rate of the plasmalemmal Ca^{2+} pump
V_{NCX}	Rate of the $\text{Na}^+/\text{Ca}^{2+}$ exchanger
V_{SERCA}	Rate of the SERCA pump
V_{RyR}	Rate of the SR Ca^{2+} release via ryanodine receptors
V_{leak}	Rate of the SR Ca^{2+} leak via ryanodine receptors

by multiplication for a log-normally distributed random scale factor k . We obtain these factors using the following MATLAB command:

$$k = \exp(\sigma * \text{randn})$$

where *randn* is a function that generates normally distributed pseudorandom numbers, and σ is the parameter that determines the degree of variability. More specifically, σ is the standard deviation of the set of scale factors k generated for each parameter (around the mean value $\mu = 1$). We use here the same value of σ (0.1) for all the parameters. Alternatively, different standard deviations can be used to reproduce the degree of variability experimentally determined for each parameter [6,7,23]. Note that while this choice does not alter the subsequent steps of this analysis, when increasing the range of variability of the model parameter, one needs to (i) discard non-physiological models; and (ii) ensure representative coverage of the parameter space (convergence of sensitivity coefficients upon stable values) by increasing the number of model variants [2]. All scale factors are allocated into the 1000×19 matrix \mathbf{K} , and used to simulate the different models in our population. Because of the relatively small σ chosen here for simplicity, APs and Ca^{2+} transients in our population fall within the normal range of experimental variability (Fig. 2), and sensitivity coefficients converge upon stable values with population sizes as small as 100 variants (not shown, see also [4]).

We repeat the proarrhythmic protocol described in [13] with our 1000 AP model variants to evaluate the occurrence of EADs in each simulation. We start all the simulations from the steady-state conditions obtained by pacing the baseline model at 1 Hz in the presence of ACh (0.1 μM). Then, we simulate instantaneous administration of isoproterenol (ISO, 1 μM) at the beginning of a 20-s phase at 10-Hz pacing, followed by 5 s at 1 Hz to mimic spontaneous termination of tachyarrhythmia (Fig. 3A). Presence of EADs is assessed during the 5 beats at 1-Hz pacing, thereby creating the array \mathbf{y} with 1000 Boolean variables (0 for absence and 1 for occurrence of at least one EAD). EADs are detected as positive deflections of the membrane potential derivative during AP repolarization.

We perform logistic regression using the MATLAB command *mnrfit*, which fits a nominal multinomial logistic regression model for the predictor matrix \mathbf{X} and the output matrix \mathbf{Y} (Fig. 1):

$$\mathbf{B} = \text{mnrfit}(\mathbf{X}, \mathbf{Y})$$

The matrix \mathbf{X} is here derived from the matrix \mathbf{K} containing the scale factors. First, \mathbf{K} is log-transformed to obtain the matrix \mathbf{K}_L :

$$\mathbf{K}_L = \log(\mathbf{K})$$

In our case, since each column in \mathbf{K} (i.e., 1000 scale factors for the same parameter) has $\mu = 1$ and $\sigma = 0.1$, each column in \mathbf{K}_L has $\mu = 0$ and $\sigma = 0.1$. To obtain \mathbf{X} , \mathbf{K}_L is converted into z-scores by subtracting

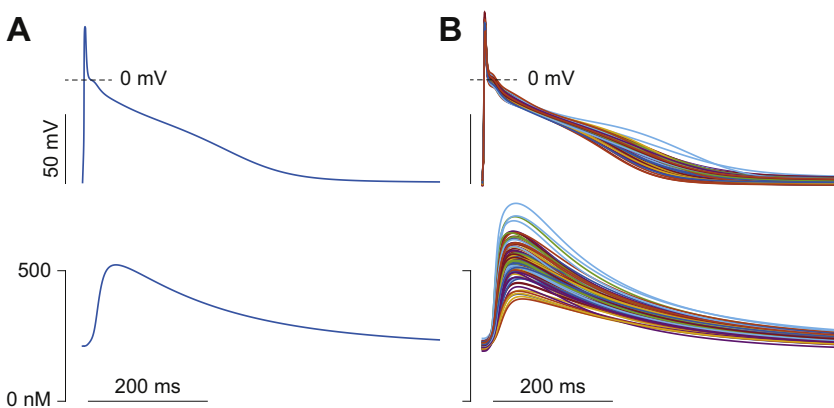


Fig. 2. Populations of AP models. AP (top panels) and Ca^{2+} transient (bottom) evoked stimulating at 1 Hz the baseline model (A) and 100 variants randomly selected in our population (B).

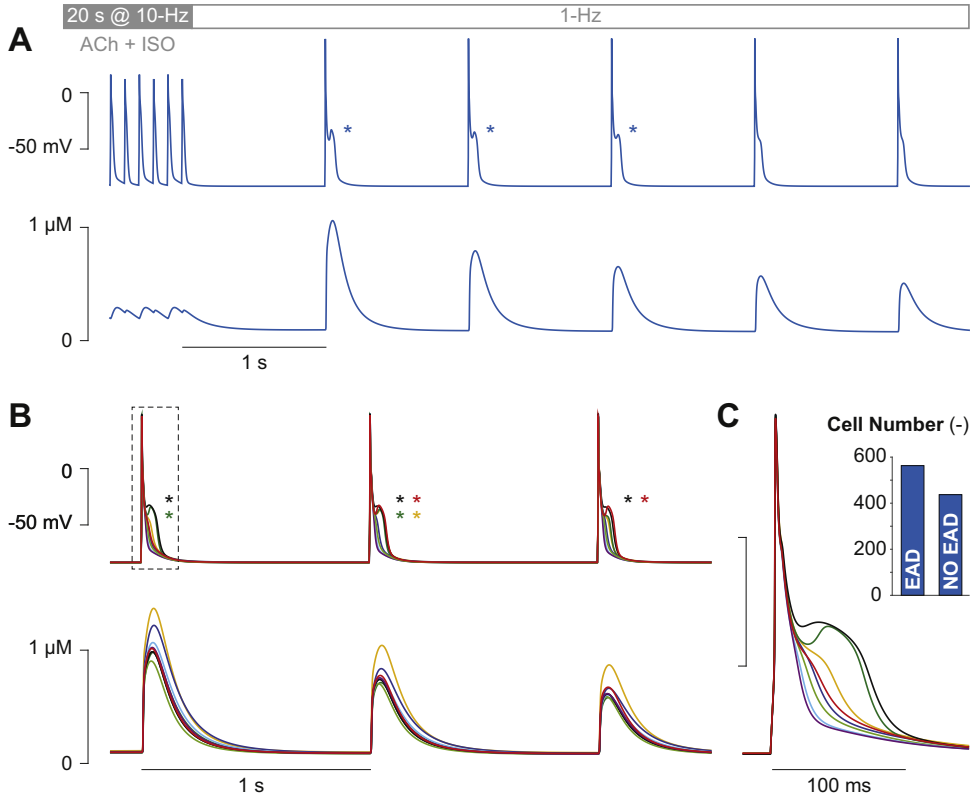


Fig. 3. Arrhythmia-provoking protocol. Membrane potential (top panels) and Ca²⁺ concentration (bottom) evoked with the baseline model (A) and 8 randomly selected variants in our population (B). EAD presence is indicated by asterisks. Panel A shows the last 6 beats of the 20-s phase at 10-Hz pacing, followed by 5 beats at normal rate (1 Hz). The first 3 beats at 1 Hz are reported in B. The first APs obtained after the pause is shown in panel C to highlight the variability among the different responses. The inset in C reports the number of simulations (i.e., models) in which at least one EAD is evoked during the 5 beats at 1-Hz pacing.

the mean across all models and dividing by the standard deviation. More specifically, we perform the following operation for each column:

$$\mathbf{x} = (\mathbf{k}_L - \mu_{kL}) / \sigma_{kL}$$

where \mathbf{k}_L and \mathbf{x} are generic columns in \mathbf{K}_L and \mathbf{X} , and μ_{kL} and σ_{kL} are mean and standard deviation computed on the column \mathbf{k}_L , respectively. Therefore, the scale factor k corresponding to \mathbf{x} , a generic element of \mathbf{X} , can be determined with the following formulation:

$$k = \exp(\sigma_{kL} \times \mathbf{x} - \mu_{kL})$$

The resulting matrix of the predictors \mathbf{X} has the same size of \mathbf{K} (1000×19) and $\mu = 0$ and $\sigma = 1$ for each column, independently of the properties of \mathbf{K} , which facilitates interpretation of the results, as described below. In fact, \mathbf{K} may contain either the scale factors (as in this example, where each column has $\mu = 1$) or the actual parameter values used in each model of the population (where the mean of each column corresponds to the parameter value in the baseline model). The degree of variation can also be different for each parameter. Since each column in \mathbf{K} may have different μ and σ values, log-transformation and following conversion into z-scores ensure that each column in \mathbf{X} has same μ and σ (0 and 1, respectively). Since here we are interested in only one output ($n_o = 1$), the matrix \mathbf{Y} is reduced to one column (1×1000), containing the binary information about presence/absence of EADs. In this case, the input of *mnrfit* must contain positive integer category numbers, thus we modify the array \mathbf{y} to

indicate the presence of EADs with 1 and absence of EADs with 2 (rather than 0). Also **B**, the output of *mnrfit*, consists of only one column (hence called array **b**) of size 1×20 , containing the intercept term b_0 and the 19 coefficients corresponding to the perturbed parameters.

When simulating the proarrhythmic protocol with our baseline model, we obtained EADs in the first 3 beats in the 1-Hz phase (Fig. 3A, see asterisks) [13]. Different behaviors are observed within the population, as shown in Fig. 3B and C. The overall analysis shows that our protocol induced EADs in 563 out of 1000 model variants (Fig. 3C, inset). In this sub-population of EAD-positive cells, only one episode (see yellow example in Fig. 3B) or multiple EADs (black example) are predicted. In 18 simulations, a delayed afterdepolarization develops between the last beat at 10-Hz and the first stimulated beat at 1-Hz (not shown). In these cases, included in the analysis, EADs are observed in 9 out of 18 models.

The 20 coefficients obtained with the logistic regression analysis are reported in Fig. 4A. The use of **X** (rather than **K**) allows easily interpreting these coefficients to assess changes in the probability that

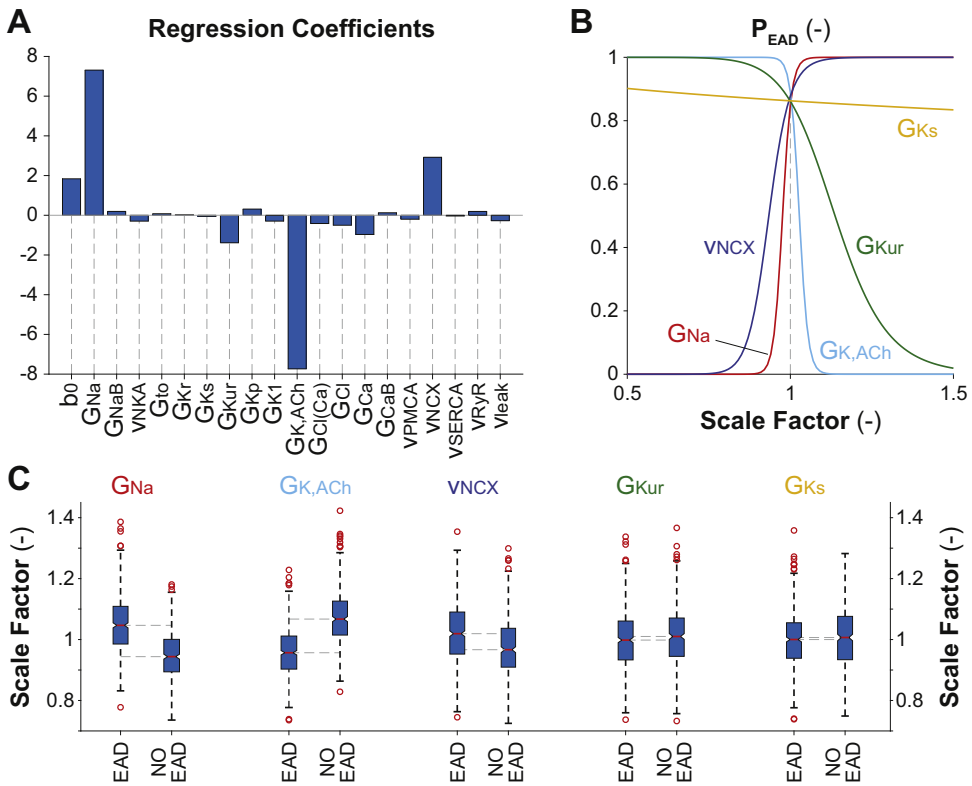


Fig. 4. Results of logistic regression analysis. A) Coefficients of the multinomial logistic regression model describing the probability of EAD development in the models of our population. B) Effects on P_{EAD} induced by modulation of one model parameter at the time (G_{Na} , $G_{K,ACH}$, V_{NCX} , G_{Kur} and G_{Ks}). C) Statistical analysis on the values of the perturbations to these 5 parameters in the two sub-populations of AP models with or without EADs performed with the MATLAB command *boxplot*. The central red line indicates the median of each group (reported in Table 2), while the blue box represents the central 50% of the data. The lower and upper boundaries of the central box are, respectively, at the 25th and 75th percentiles. The dotted vertical lines (called “whiskers”) maximally extend to 1.5 times the height (called “inter-quartile range”) of the central box. Individual values outside the range of the whiskers are considered outliers (shown here with red circles). The extremes of the lateral notches of the central box are determined as $q_{50} \pm 1.57(q_{75} - q_{25})/\sqrt{n}$, where q_{25} , q_{50} and q_{75} are the 25th, 50th (i.e., the median) and 75th percentiles, respectively, and n is the number of observations in each group. These values (reported in Table 2) mark the 95% confidence interval for the medians (i.e., if the notches from two boxplots do not overlap, one can assume that the medians are different with a significance level of 0.05).

the cell will develop EADs (P_{EAD}). Except for b_0 , the values of all the other regression coefficients reflect the effect of 1-standard deviation increase from the (log-transformed) baseline value of the corresponding model parameter. Hence, all these coefficients can be directly compared, even if the related parameters are characterized by different variability ranges around different mean values [2]. Augmenting the parameters associated with positive coefficients increases P_{EAD} . This probability increases also when reducing the parameters associated with negative coefficients. The largest bars correspond to model parameters that mostly influence the development of EADs, namely G_{Na} , $G_{K,ACh}$, v_{NCX} , and G_{Kur} . In Fig. 4B we quantify the effect of modulation in these parameters (and in G_{Ks}) on P_{EAD} . These traces are obtained by assuming changes in one parameter at the time (with respect to the baseline value), using the following equation:

$$P_{EAD}(x, b) = 1/(1 + \exp(-(b_0 + x*b)))$$

where b is the regression coefficient corresponding to the perturbed parameter, and x is the degree of variation (log-transformed and converted into z-score). Here we show the effects of $\pm 50\%$ modulation in these model parameters (i.e., corresponding to variations of the scale factors between 0.5 and 1.5). Note that, given a scale factor for a certain parameter, the corresponding x is obtained using the values of μ_{KL} and σ_{KL} estimated for the same parameter in the population, as described before. In absence of perturbations (i.e., when $x = 0$), P_{EAD} depends only on b_0 . In this case, which corresponds also to the probability calculated for the baseline model (where the array \mathbf{x} is null), P_{EAD} is high (~ 0.86). Indeed, we obtain EADs in this condition (Fig. 3A).

As described in our previous paper, these phase-3 EADs are a consequence of non-equilibrium reactivation of I_{Na} , and their occurrence can be prevented by blocking I_{Na} with ranolazine [13]. In our preliminary sensitivity analysis (see Supplementary data in [13]) we also showed that the same result can be obtained with 5% reduction in G_{Na} , but not with $\pm 5\%$ modulation in other model parameters. Thus, finding that G_{Na} is the most important parameter (largest bar in Fig. 4A, and steepest relationship in Fig. 4B) is not unexpected. Fig. 4B shows that also changes in $G_{K,ACh}$ and v_{NCX} have important and opposite effects on P_{EAD} : increased $G_{K,ACh}$ (negative coefficient) and decreased NCX activity (positive coefficient) are protective. The involvement of NCX is also to be expected. Indeed, our proposed mechanism involves I_{Na} reactivation, which is favored by NCX-mediated AP plateau prolongation. NCX enhancement is induced by Ca^{2+} augmentation upon re-establishment of normal pacing frequency. Decreased v_{NCX} shortens the AP, thereby hindering I_{Na} reactivation. The same effect is also obtained when the repolarizing ACh-sensitive and ultra-rapid K^+ currents are enhanced (i.e., increased $G_{K,ACh}$ and G_{Kur}). Thus, while $I_{K,ACh}$ -induced APD shortening is required to hasten I_{Na} recovery from inactivation, excessive $I_{K,ACh}$ prevents EADs by shifting the balance between repolarizing and depolarizing currents. Indeed, we had previously observed that increasing [ACh] prevents EADs in the model, primarily by increasing the repolarization rate [13].

We compare the scale factor values relative to these parameters in the two EAD-positive and EAD-negative sub-populations. Fig. 4C shows the results of a statistical analysis performed (and plotted) with the MATLAB command *boxplot*. The distributions of the scale factors of the most influential

Table 2

Properties of the scale factor distributions in the two sub-groups.

Parameter	EAD	Median		Notches (low/high)	Mean	Standard deviation
G_{Na}	Yes	1.0466	*	1.0384/1.0548	1.0519	0.0948
	No	0.9439		0.9359/0.9519	0.9482	0.0815
$G_{K,ACh}$	Yes	0.9567	*	0.9495/0.9639	0.9589	0.0803
	No	1.0672		1.0588/1.0756	1.0723	0.0906
v_{NCX}	Yes	1.0193	*	1.0102/1.0284	1.0206	0.1007
	No	0.9667		0.9571/0.9763	0.9744	0.0948
G_{Kur}	Yes	0.9982		0.9898/1.0067	0.9991	0.0962
	No	1.0101		1.0006/1.0195	1.0100	0.0990
G_{Ks}	Yes	1.0002		0.9925/1.0079	1.0012	0.0902
	No	1.0067		0.9960/1.0174	1.0100	0.0989

*Asterisks indicate “strong evidence” (95% confidence) that the medians differ [25].

parameters (G_{Na} , $G_{K,ACh}$, and v_{NCX}) are markedly different in the two sub-groups (Fig. 4C). This simple analysis shows that the median values are different with a significance level of 0.05 (Fig. 4C and Table 2). Instead, there are no differences between the median values for G_{Kur} , despite this parameter is shown to be important for P_{EAD} (see Fig. 4A). For completeness, we also report an example relative to a parameter that does not play a significant role in EAD formation in this context, such as G_{Ks} , i.e., the maximal conductance of the slowly activating K^+ current. The low regression coefficient associated to this parameter (Fig. 4A) makes P_{EAD} insensitive to perturbations in G_{Ks} (see flat relationship in Fig. 4B). Accordingly, also the distributions of the scale factors associated to G_{Ks} in the two sub-populations are not different (Fig. 4C and Table 2).

Conclusions

Multivariable logistic regression analysis, applied here to investigate the factors responsible for arrhythmia development, is suitable for studying any process characterized by dichotomous outcome in both deterministic and stochastic models of myocyte electrophysiology and Ca^{2+} handling. For example, this predictive method has been applied successfully to analyze a stochastic model of Ca^{2+} sparks in the ventricular cell [2], and to investigate the drug-induced loss of automaticity in a deterministic model of the sinoatrial myocyte [8]. By using receiver operator characteristic analysis, Lee et al. [2] showed that logistic regression has the same predictive power of support vector machine, an established method for classifying binary data [24]. They also determined that this methodology is computationally more efficient than one-at-a-time parameter sensitivity analysis [2]. Therefore, application of logistic regression to populations of mathematical models is an efficient and effective technique that allows gaining novel biological insight and generating experimentally testable hypotheses, thereby improving our understanding of cardiac function in health and disease.

Acknowledgements

We would like to thank Dr. Eric Sobie, who suggested applying this logistic regression method to our EAD studies, and Nicholas Ellinwood for critical reading of the manuscript. This work was supported by the National Institutes of Health R01 Grant HL131517 (EG), the American Heart Association Scientist Development Grant 15SDG24910015(EG), and the Heart Rhythm Society post-doctoral fellowship160A9HRS (SM).

References

- [1] E.A. Sobie, Parameter sensitivity analysis in electrophysiological models using multivariable regression, *Biophys. J.* 96 (4) (2009) 1264–1274.
- [2] Y.S. Lee, O.Z. Liu, H.S. Hwang, B.C. Knollmann, E.A. Sobie, Parameter sensitivity analysis of stochastic models provides insights into cardiac calcium sparks, *Biophys. J.* 104 (5) (2013) 1142–1150.
- [3] M.A. Cummins, P.J. Dalal, M. Bugana, S. Severi, E.A. Sobie, Comprehensive analyses of ventricular myocyte models identify targets exhibiting favorable rate dependence, *PLoS Comput. Biol.* 10 (3) (2014) e1003543.
- [4] R.A. Devenyi, E.A. Sobie, There and back again: iterating between population-based modeling and experiments reveals surprising regulation of calcium transients in rat cardiac myocytes, *J. Mol. Cell. Cardiol.* 96 (2016) 38–48.
- [5] L. Romero, E. Pueyo, M. Fink, B. Rodriguez, Impact of ionic current variability on human ventricular cellular electrophysiology, *Am. J. Physiol. Heart Circ. Physiol.* 297 (4) (2009) H1436–45.
- [6] O.J. Britton, A. Bueno-Orovio, K. Van Ammel, H.R. Lu, R. Towart, D.J. Gallacher, et al., Experimentally calibrated population of models predicts and explains intersubject variability in cardiac cellular electrophysiology, *Proc. Natl. Acad. Sci. U. S. A.* 110 (23) (2013) E2098–105.
- [7] E. Passini, A. Mincholé, R. Coppini, E. Cerbai, B. Rodriguez, S. Severi, et al., Mechanisms of pro-arrhythmic abnormalities in ventricular repolarisation and anti-arrhythmic therapies in human hypertrophic cardiomyopathy, *J. Mol. Cell. Cardiol.* 96 (2016) 72–81.
- [8] S. Morotti, R. Clair St., R. Proenza, E. Grandi, Dynamic regulation of pacemaker activity by the $Na^+ - K^+$ pump, *Comput. Cardiol.* 43 (2016), doi:<http://dx.doi.org/10.22489/CinC.2016.221-298>.
- [9] A.X. Sarkar, D.J. Christini, E.A. Sobie, Exploiting mathematical models to illuminate electrophysiological variability between individuals, *J. Physiol.* 590 (11) (2012) 2555–2567.
- [10] R.H. Johnstone, E.T. Chang, R. Bardenet, T.P. de Boer, D.J. Gavaghan, P. Pathmanathan, et al., Uncertainty and variability in models of the cardiac action potential: can we build trustworthy models, *J. Mol. Cell. Cardiol.* 96 (2016) 49–62.
- [11] N.R. Draper, H. Smith, *Applied Regression Analysis*, John Wiley & Sons, 2014.
- [12] J.F. Hair, *Multivariate Data Analysis*, Prentice Hall, 2010.

- [13] S. Morotti, A.D. McCulloch, D.M. Bers, A.G. Edwards, E. Grandi, Atrial-selective targeting of arrhythmogenic phase-3 early afterdepolarizations in human myocytes, *J. Mol. Cell. Cardiol.* 96 (2016) 63–71.
- [14] E. Grandi, S.V. Pandit, N. Voigt, A.J. Workman, D. Dobrev, J. Jalife, et al., Human atrial action potential and Ca²⁺ model: sinus rhythm and chronic atrial fibrillation, *Circ. Res.* 109 (9) (2011) 1055–1066.
- [15] A. Burashnikov, C. Antzelevitch, Reinduction of atrial fibrillation immediately after termination of the arrhythmia is mediated by late phase 3 early afterdepolarization-induced triggered activity, *Circulation* 107 (18) (2003) 2355–2360.
- [16] E. Patterson, R. Lazzara, B. Szabo, H. Liu, D. Tang, Y.H. Li, et al., Sodium-calcium exchange initiated by the Ca²⁺ transient: an arrhythmia trigger within pulmonary veins, *J. Am. Coll. Cardiol.* 47 (6) (2006) 1196–1206.
- [17] E. Patterson, S.S. Po, B.J. Scherlag, R. Lazzara, Triggered firing in pulmonary veins initiated by in vitro autonomic nerve stimulation, *Heart Rhythm* 2 (6) (2005) 624–631.
- [18] D. Amar, H. Zhang, S. Miodownik, A.H. Kadish, Competing autonomic mechanisms precede the onset of postoperative atrial fibrillation, *J. Am. Coll. Cardiol.* 42 (7) (2003) 1262–1268.
- [19] M. Bettoni, M. Zimmermann, Autonomic tone variations before the onset of paroxysmal atrial fibrillation, *Circulation* 105 (23) (2002) 2753–2759.
- [20] T. Tomita, M. Takei, Y. Saikawa, T. Hanaoka, S. Uchikawa, H. Tsutsui, et al., Role of autonomic tone in the initiation and termination of paroxysmal atrial fibrillation in patients without structural heart disease, *J. Cardiovasc. Electrophysiol.* 14 (6) (2003) 559–564.
- [21] N. Voigt, J. Heijman, Q. Wang, D.Y. Chiang, N. Li, M. Karck, et al., Cellular and molecular mechanisms of atrial arrhythmogenesis in patients with paroxysmal atrial fibrillation, *Circulation* 129 (2) (2014) 145–156.
- [22] A.G. Edwards, E. Grandi, J.E. Hake, S. Patel, P. Li, S. Miyamoto, et al., Nonequilibrium reactivation of Na⁺ current drives early afterdepolarizations in mouse ventricle, *Circ. Arrhythm. Electrophysiol.* 7 (6) (2014) 1205–1213.
- [23] A. Muszkiewicz, O.J. Britton, P. Gemmell, E. Passini, C. Sanchez, X. Zhou, et al., Variability in cardiac electrophysiology: using experimentally-calibrated populations of models to move beyond the single virtual physiological human paradigm, *Prog. Biophys. Mol. Biol.* 120 (1–3) (2016) 115–127.
- [24] A. Ben-Hur, C.S. Ong, S. Sonnenburg, B. Scholkopf, G. Ratsch, Support vector machines and kernels for computational biology, *PLoS Comput. Biol.* 4 (10) (2008) e1000173.
- [25] J.M. Chambers, W.S. Cleveland, B. Kleiner, P.A. Tukey, Comparing data distributions, *Graphical Methods for Data Analysis*, Belmont, California, Wadsworth International Group, 1983.